DOCUMENT RESUME

ED 051 297 TM 000 618

AUTHOR	Badal, Alden W.; Larsen, Edwin P.
TITLE	Measurement in Education: On Reporting Test Results to Community Groups.
INSTITUTION	National Council on Measurement in Education, East Lansing, Mich.
PUB DATE	May 70
NOTE	12p.; Special Report, v1 n4 1970
AVAILABLE FROM	National Council on Measurement in Education, Office of Evaluation Services, Michigan State University, East Lansing, Michigan 48823. $2.00 per year (4 issues); Single issues 0.25 each in quantities of 25 or more
EDRS PRICE	EDRS Price MF-$0.65 HC Not Available from EDRS.
DESCRIPTORS	Achievement Tests, Audiovisual Aids, Grade Equivalent Scores, *Information Dissemination, Measurement, Models, Norms, Parent School Relationship, *Public Relations, School Community Relationship, *Scores, Student Grouping, Test Bias, Testing Programs, *Test Interpretation, *Test Results

ABSTRACT

The major elements of a test interpretation model which would assist school personnel in presenting standardized test information to the public are presented. The model is a prototype based upon the testing program used in the Oakland, California Public Schools. An outline and sequence of the test score presentation are suggested, including notes on important background concepts. A discussion of test scores as they reflect school needs, and a selection of questions frequently asked by parent and community groups are provided. Consideration is given to uses of test scores, questions in interpreting test results, types of tests given in schools, test norms on comparison groups, types of test scores, summary statistics, and suggested data for presentation. Statistical illustrations are provided. (AE)

ED051297

Volume 1, No. 4

May, 1970

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

NCME

measurement in education

A SERIES OF SPECIAL REPORTS OF THE NATIONAL COUNCIL ON MEASUREMENT IN EDUCATION

On Reporting Test Results to Community Groups

Alden W. Badal

Edwin P. Larsen

ABOUT THIS REPORT

Measurement specialists usually devote their energies to developing new ways to talk to one another. But educational measurement is not only applied in situations where technical language is necessary to communicate professional understanding. Increasingly, the techniques, implications, and the effects of educational measurement are coming under scrutiny as the public comes to understand that testing is a critical form of social evaluation. Consequently, it becomes more and more important to develop a variety of ways to talk about testing to interested persons at varying levels of technical sophistication.

This report provides a down-to-earth set of guidelines which school personnel can use in developing presentations to lay audiences. A variety of basic terms and ideas are enumerated and described in language which should prove helpful to those facing the task of explaining a testing program.

In providing this valuable service, the authors speak from solid experience. Dr. Badal reached his present post as Director of Research for the Oakland District through a variety of teaching, research and administrative positions. He serves on a number of professional committees and boards. Mr. Larsen is Supervisor of Research and Group Testing in Oakland and has been active in statewide testing in California.

WWW

Alden W. Badal & Edwin P. Larsen

The public release of school-by-school test averages, while not a common practice as yet, is becoming a more frequent occurence, especially in some of the larger urban school districts. New York, San Francisco, Detroit, Philadelphia, Baltimore, Los Angeles, and Sacramento are examples of city districts that have adopted this procedure during recent years. The decision to take this step was made also in Oakland, and the first public school-by-school test report was made to the Oakland Board of Education in December of 1967.

Experiences reported have revealed that school administrators in all of these districts agree that a major challenge associated with this practice is that of developing adequate public understanding of the meaning of the various types of test scores, statistics, and terminology which are used so that misinterpretations and incorrect conclusions can be minimized. As in many districts, the Research Department of the Oakland Public Schools has attempted, over the years, to provide assistance upon request to school principals in the presentation of test score data to faculties and to school-community organizations and groups. However, at the time of the decision to release school test score averages to the public it was anticipated that requests for assistance with such test interpretation meetings might increase to the point that the Research Department would be unable to meet all or even a major portion of the requests.

Consequently, it was concluded that it would be necessary for school principals to assume greater responsibilities in presenting test data for their own schools to local community groups. It was for this reason that the Research Department developed a model presentation entitled *Tentative Guidelines for Interpretation of Test Scores* which outlines in detail a number of important test interpretation procedures. These "Tentative Guidelines" also included a companion series of twenty-one overhead projector transparencies for use during the test interpretation presentations.

T: 000 618

It is the purpose of this report to present the major elements of this test interpretation model with the hope that this information will be of assistance to school personnel elsewhere in developing improved techniques and procedures for presenting test information to the public. Those familiar with testing will find many of the concepts and much of the information presented quite basic and frequently stated in very elementary terms. This is intentional since the audience for whom the test presentations are intended generally has only limited knowledge about standardized testing.

The test interpretation model presented is a prototype based upon the testing program in the Oakland Public Schools. Obviously, it may include features that are not characteristic of many testing programs; on the other hand, it may omit features which are important components of testing situations elsewhere.

The suggested test interpretation presentation takes approximately an hour to complete when time is allowed for questions and discussion. Some sections in the "Tentative Guidelines" statement can be omitted or dealt with only briefly if, in the principal's judgment, this would be appropriate for a particular audience. However, it is emphasized that most parents and some teachers have only limited understanding of the meaning of terms such as grade placements, percentiles, medians, quartiles, and state norms. Therefore, it is recommended that the terms used in the presentation be carefully explained before attempting to present analyses of test score data.

As stated previously, a series of overhead projector transparencies were developed for use in the presentation. Such visual aids can be extremely useful in presenting technical material to lay audiences, and it has been found that many ideas which are necessary to communicate are readily grasped with charts or figures. Consequently, this report includes examples of several visuals which have proven useful in the Oakland Public Schools. Others can be easily constructed depending upon available test results and the types of problems which deserve attention.

The "Tentative Guidelines" statement was designed to provide suggestions and information in these three areas:

- A suggested outline and sequence of the test score presentation including notes on important background concepts to be developed in the presentation.
- A discussion of test scores as they reflect school needs.
- A number of selected questions frequently asked by parent and community groups.

In order to provide as clear an account as possible concerning the guidelines provided in these three areas, the remaining portion of this report is in large part a presentation of excerpts from the "Tentative Guidelines" statement accompanied by examples of some of the charts and figures that have been found to be particularly useful. It will be noted that the organization and content have been prepared in such a way as to provide the principal with certain needed factual information, a specific series of steps to be followed if desired, and simply stated language and terminology which may be used in the interpretation session. The purpose here is not to discuss technical aspects of test interpretation, but to communicate with an audience that may frequently be uninformed. Much of the following material is intended to suggest possibilities which might be adapted to local circumstances and the intended audience.

SUGGESTED OUTLINE AND SEQUENCE

I. The Testing Program in the Oakland Public Schools

While the majority of the discussion of test results may focus on the State Testing Programs, it is important to develop an awareness that testing has been a regular part of the local educational program for many years. Many teachers use reading readiness tests at the kindergarten or early first grade level to assist them in planning for the most appropriate program in first grade. From that point on through the grades, children in the Oakland schools are tested to obtain information for a number of purposes.

Less than 1% of the total instructional time is devoted to testing at any grade level. We feel that this is not too great an amount of time to devote to testing because of the contributions of test scores to the measurement of the progress of individuals and the effectiveness of our instructional program. (It should be pointed out here that test scores are certainly not the only type of information used in evaluating a school program.)

Some of the most important uses of test scores are these:

A. Evaluation of the growth of individual children. Periodic measures in basic skills (usually reading, writing, mathematics skills) on pupils as they progress through the grades is important information for the school to have.

B. Diagnosis of strengths and weaknesses. An example is the growth of a child in reading as compared with mathematics or writing skills.

C. Grouping students for instruction. There are sizable ranges of ability or skills among the chil-

dren within a school. While we do not subscribe to "tracking," or grouping strictly by ability, it is recognized that all children may not profit from a single lesson or the same book. For some, a lesson may be too difficult; for others, much too easy. Since children from all ability levels will learn best from lessons and materials which are geared to their present levels of development, principals usually try to place two or three clusters or groups of children within a classroom. Usually, there will be children of below average, average, and above average abilities in any class. Test scores are one type of information which helps in grouping children for instruction.

D. Evaluation of instruction. By examining score averages for groups of children it is possible to study which types of learning materials or teaching methods are most effective. For example, the Oakland Public Schools cooperated in a study of the effectiveness of four different types of reading programs at the primary grade levels. Test scores, along with other types of evaluation information, were used in this study. The School Mathematics Study Group has also been working with different types of mathematics materials in our schools, and test scores have played an important part in determining the effectiveness of these different materials. And finally, in almost all the federally funded compensatory education programs, school districts are required to administer various kinds of achievement tests to help determine the effectiveness of the various services and programs.

E. Informing the general public and parents. We feel that parents have a right to know all that the school knows about the educational strengths and weaknesses of their children. But we do feel that test scores must be explained so that parents understand the meanings of the special terms which are used in discussing test results.

II. Understanding Test Scores

A. Three Key Questions in Interpreting Test Results.

There are three key questions which should be kept in mind when studying the scores of a single student or the score averages of a group of students (See Figure 1).

What skills are being measured or tested? For example, just what does a student need to know in order to get a high score on a certain reading test?

With whom are pupils being compared? All test scores are basically a comparison of a pupil's performance with some other pupils' – usually a group of the same age. It is important to keep in mind who these "other" pupils are.

1. What skills are measured?

- Reading
- Arithmetic
- Language skills
- Spelling
- Scholastic aptitude (intelligence)

2. With whom are pupils being compared?

- Typical pupils across the nation (National Norms)
- Pupils throughout the state (State Norms)
- Pupils throughout the city (Local Norms)

3. What type of score is being used?

- Grade placements
- Percentile ranks

Figure 1. Understanding Test Scores.

What type of score is being used? Pupils may be compared using a number of different types of scores. What do grade equivalents and percentile scores mean?

In helping parents understand test scores, as much time may be spent in attempting to develop background concepts as in talking about the actual test score statistics.

B. Types of Tests Given in the Schools. There are two main types of tests which are given to children in the elementary schools:

Achievement tests. These tests are made up of questions on types of materials and understandings which are commonly taught in the schools throughout the country. For example, in a reading test children are asked to read short paragraphs similar to those found in textbooks and to answer questions to show how well they have understood what has been read. Arithmetic or mathematics tests contain addition, subtraction, multiplication and division problems. Also, they often contain questions to show how well students can solve word problems. A writing test usually asks pupils to determine whether or not a sentence has proper punctuation, capital letters, and spelling. Pupils are also asked to select sentences which express an idea most clearly.

Scholastic aptitude tests. This is the second type of test most often administered in the schools. These tests are sometimes loosely referred to as intelligence or "IQ" tests. The test makers themselves perfer to consider these tests as measuring school abilities. Tests of this type are intended to give an estimate of the level of achievement test performance which may be "expected" of a given pupil or group of pupils. These scholastic aptitude tests attempt to measure a student's background of understandings which are learned from both in-school and out-of-school experiences. Scholastic aptitude tests are more general in nature than achievement tests, but they must not be considered to measure innate or unchanging aspects of "intelligence." They merely give a reasonable estimate of the background or tool skills of a pupil at the present time and, therefore, how well we may expect him to perform in reading, reason with numbers, and use the English language in written work.

Staff members of the Oakland Public Schools recognize that considerable concern has been registered throughout the country over the appropriateness of utilizing scholastic ability tests, particularly with minority children. However, to date the district has operated under the assumption that the advantages of using these instruments have outweighed the disadvantages and has continued to administer and utilize these tests at selected grade levels.

C. Test Norms or Comparison Groups. The most common types of standardized test scores are those which compare a pupil's performance with a "national norms group." The publisher of a test carefully selects a sample of children from all parts of the nation in order to estimate what is the average or normal level of performance of children at a given grade level. This procedure is much like that used by opinion polls. The information obtained from this sample becomes the publisher's "national norms," and we can determine how a child or group of children compares to this norm. Since these norms are based on only a sample of children, it is occasionally found that these norms are not completely accurate. They may be "too easy" or "too hard."

In California at certain grade levels a pupil's scores can also be compared to other children in the same grade throughout the state. The state score patterns and averages based on testing all children in the state at selected grade levels are used as "state norms." We can tell, for example, how a sixth grader ranks in reading ability in comparison to other sixth graders throughout the state.

The score averages of the city provide yet another comparison in evaluating local school results. The question here is, "How well does a pupil or school rank within the city?"

One can see that it is important to keep in mind the group with which pupils are being compared.

D. Types of Scores. We have discussed the skills which the tests measure and the group with whom pupils are being compared. The next question is what type of score is used to make these comparisons. There are two types of scores which we will use in our discussion.

Grade equivalent scores. Grade equivalents, sometimes called grade placement or just grade scores, are often used because the number of correct answers on a test has little meaning in itself. Publishers administer the test to children in dif-

NCME

measurement in education

Volume 1, No. 4 May, 1970

Warren W. Willingham, Editor

Advisory Committee:
J. Thomas Hastings, University of Illinois
Alice J. Irby, Educational Testing Service
Robert S. Lankton, Detroit Public Schools

Measurement in Education is a series of special reports published four times a year in October, January, March, and May by the National Council on Measurement in Education. These reports are concerned with the practical implications of measurement and related research and their application to educational problems of individuals, institutions, and systems. The emphasis is upon uses of measurement rather than technical or theoretical issues. Subscription rate: $2.00 per year; single copies 30¢ each in quantities of 25 or more.

Second class postage paid at East Lansing, Michigan 48823

Address Editorial Correspondence to:

Warren W. Willingham
College Board
800 Welch Road
Palo Alto, Ca. 94304

Address Business Correspondence to:

Irvin J. Lehmann
Office of Evaluation Service
Michigan State University
East Lansing, Mich. 48823

ferent grade levels during a given school month. The average of a grade level is assigned a grade equivalent value such as 1.9, 2.9, 3.9 (See Figure 2). With these grade equivalent values determined, the test authors then estimate the in-between values. We can, therefore, compare a pupil's raw score with other children. One will note scores such as 1.3, 5.6, and so on. In Figure 2 "1.3" refers to the score earned by an "average" first grader at the 3rd month of the school year; and "5.6" means a score level which is typical at the 6th month of grade five.

Percentiles. Percentiles are a type of test score which is also useful in comparing the test performance of children. A percentile is not the same as percent correct which is often seen on such teacher-made tests as spelling or arithmetic. A percentile is a pupil's rank within a theoretical group of 100 pupils who represent the patterns of ability or skill across the country, school district, or other normative group.

Students differ widely in their abilities and performance. If we were to select a hundred "typical" youngsters from all parts of the nation and ask them to run a one-mile race, we would soon find differences in running skills. Figure 3 indicates the possible results of that race. Let us say, for example, that almost all children run more rapidly than the "orange" boy. The "freckled" boy is fairly average; about half of the children are ahead and half are behind. And very few are as fast as the "gray" boy. We can rank these children using the percentile system.

Number Correct	Grade Equivalent	Interpretation
20	4.3	
19	4.1	Score earned by "average" 3rd grader in ninth month of grade 3.
18	3.9	
17	3.7	
16	3.5	
15	3.3	
14	3.1	Score earned by "average" 2nd grader in ninth month of grade 2.
13	2.9	
12	2.7	
11	2.5	
10	2.3	
9	2.1	Score earned by "average" 1st grader in ninth month of grade 1.
8	1.9	
7	1.8	
6	1.7	
5	1.6	
4	1.5	
3	1.4	
2	1.3	
1	1.2	
0	1.1	

Figure 2. Interpretation of Grade Equivalent Scores.

In Figure 3 the orange boy ranks at the 4th percentile – 4 out of 100 run as fast or more slowly than this boy; 96 out of 100 are faster. The freckled boy falls in the middle area where most children fall with his percentile rank of 44. This means that 44 out of every one hundred pupils run as fast or slower; 56% run faster. Finally, the gray

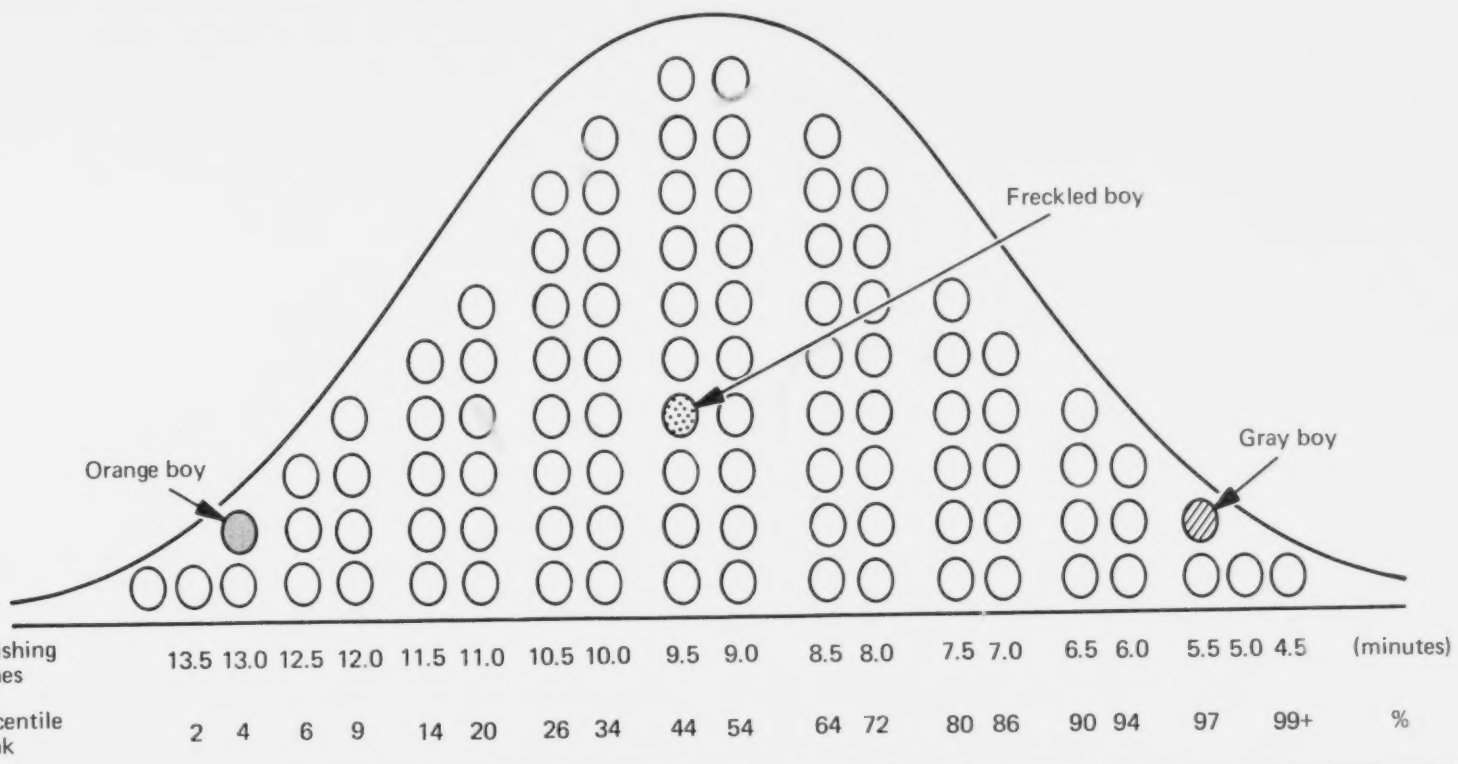


Figure 3. Interpretation of Percentile Scores.

boy receives a score at the 97th percentile, meaning that 97 out of one hundred run this fast or slower; only 3 out of one hundred run more rapidly. This same ranking system is used to describe scholastic ability and achievement test scores. Percentiles tell how many children out of every one hundred in the state or nation score at or below a certain score level. By subtracting a percentile value from 100, the number scoring higher can be determined.

The 50th percentile is considered to be right "on the grade level" or right at the average of the norm group. Pupils may vary somewhat either above or below this value and still be considered within the "average range." In general the percentile range of 25 through 75 is considered to be the average range so that pupils who are a few points above or below the 50th percentile are considered to be essentially average. Students above and below the 25 to 75 range are clearly above or below average.

E. Summary Statistics. When tests are administered to groups of pupils, there will be quite a range of scores – some pupils may get nearly all questions correct; others, very few correct. The distribution of scores on a reading test shown in Figure 4 will illustrate this fact. In order to summarize this array of scores, median and quartile statistics are used.

Number Correct	Number of Students	
20	I	
19	II	
18	I	
17	III	
16	II	
15	⧵⧵⧵⧵ I	
14	⧵⧵⧵⧵	
13	⧵⧵⧵⧵ III	Q_3 — ¾ or 75% at or below a score of 13.
12	⧵⧵⧵⧵ ⧵⧵⧵⧵ II	
11	⧵⧵⧵⧵ ⧵⧵⧵⧵ I	Mdn — ½ or 50% at or below a score of 11.
10	⧵⧵⧵⧵ III	
9	⧵⧵⧵⧵ ⧵⧵⧵⧵	
8	⧵⧵⧵⧵	Q_1 — ¼ or 25% at or below a score of 8.
7	⧵⧵⧵⧵ II	
6	⧵⧵⧵⧵	
5	III	
4	III	
3	I	
2	II	
1		
0	I	
Total Students	96	

Figure 4. Distribution of Reading Test Scores for School "X".

- Q_1 indicates the score at or below which ¼ or 25% of the pupils fell. Three fourths of the pupils scored higher.

- The median indicates the middle point in the score range. One half of the students are above and one half are below this point. This is important to keep in mind. A median obviously does not reflect the scores of all children – many pupils are somewhat higher; many are somewhat lower.

- Q_3 indicates the score at or below which ¾ or 75% of the children have scored. One fourth have scored higher.

If the size of a school is farily small – 50 or fewer pupils – caution must be used in interpreting medians and quartiles. A change in the score of one pupil may cause the average to change several points.

III. Suggested Data for Presentation

The Oakland School District selects test results at various grade levels at different time intervals in order to present information which provides the most useful and meaningful interpretation to parents. Frequently, a determining factor is the test score and normative information available from local and statewide programs in particular years. Some important types of comparisons which this district has found helpful are as follows:

A. National-State-District-School. A question in which most audiences are immediately interested is how the test results look from a normative standpoint. How does the state compare with the whole country? Is the district up to the state average? Is this school above or below average for the district? Figure 5 shows the grade equivalents for a reading test at one grade level for these four normative groups. It is generally necessary to comment on the extent of the differences. Any that are large enough to be educationally significant might be discussed. Sometimes there are simple reasons for differences between school and district scores, for example, which do not occur to parents. Figure 5 can be produced with percentiles based upon state data. This can give a more complete picture of district and individual school scores.

B. Year to Year. An examination of the test results from preceding years can be informative. Marked changes from one year to another are not typical, and consistent results shown for two years as in Figure 5 may lend credence to current results. Real changes are more likely to occur gradually and are best portrayed in a chart such as that shown in Figure 6. Long-range trends can usually be attributed to changes in the socioeconomic

6

	Norms Group	State		District		School	
	1963	May 1966	May 1967	May 1966	May 1967	May 1966	May 1967
				Grade 1			
Q_3	2.3	1.8	1.9	1.8	1.9	1.7	1.7
Mdn	1.9	1.6	1.6	1.6	1.6	1.6	1.5
Q_1	1.7	1.4	1.5	1.4	1.5	1.4	1.4
				Grade 2			
Q_3	3.4	3.0	3.1	3.0	2.9	2.7	2.9
Mdn	2.9	2.5	2.5	2.3	2.4	2.1	2.3
Q_1	2.4	1.9	1.9	1.8	1.9	1.7	1.8

*Publisher's national grade equivalent norms

Figure 5. Stanford Reading Test Results* for Two Grades in Two Years.

characteristics of the community or in the school curriculum. However, other possibilities should be considered as well.

C. Grade by Grade. Additional tables such as the one shown in Figure 5 can be used to show the results for one test through successive grades. A set of such tables can demonstrate, for example, that a slight downward or upward trend in reading achievement from grade one to grade three in one school may also be characteristic of the whole district. Or it might be demonstrated that average mathematics skills scores in a school or district remain at grade level throughout the elementary school years despite a slight progressive loss for the state as a whole.

D. Achievement Versus Scholastic Ability. Figure 7 shows a convenient method for displaying the results of scholastic ability and achievement tests. A particular advantage of this combination of ability and achievement data is the fact that it encourages a more realistic view of the performance of pupils in a given school or district. Thus, high achievement in a school may be partly explained by high aptitude; or conversely, below average achievement may be a reflection of relatively low aptitude.

E. Subject by Subject. Variations on Figure 7 can include scores for a variety of different subject matter areas in order to demonstrate strengths and weaknesses. Such comparisons among subject areas are common and useful. In mathematics, areas of primary concern are concepts, skills, and application; in reading there is special interest in the relative strength in speed, comprehension, and vocabulary. Parents are particularly concerned with these comparisons and can easily see the possible relationship between relative score levels and the school curriculum. Discussion of subject-by-subject comparisons or skills within subjects provides a good opportunity to bring out positive action the school is taking in order to help pupils achieve better. It is also an opportunity to discuss various needs of the school (see section following).

F. Comparison of Distributions. It is important to point out that median and quartile statistics only describe the score patterns in a general way. Figure 8 provides a closer look at the spread of scores for a set of related tests. The line labeled "Norms %" indicates the proportion of students who would be expected to score at various levels – either on the basis of national, statewide, or district norms. The additional lines provide space to fill in actual numbers of students who score at various levels on different tests in a particular school. That is, if a school had "normal" numbers of pupils at each of

CITY MEDIANS – GRADE 5

Year	STEP Mathematics	STEP Reading	STEP Writing	SCAT Verbal	SCAT Quantitative	SCAT Total
67	44	41	40	44	43	43
66	41	43	43	39	36	40
65	41	43	40	39	41	40
64	48	50	47	46	46	45
63	44	41	41	42	46	45
62	47	45	43	41	43	43

Figure 6. Median Percentile Results from SCAT and STEP in Successive Years.

	School Ability (Lorge-Thorndike I.Q.)				CTBS Reading (Grade Equivalent)				CTBS Arithmetic (Grade Equivalent)		
	Norms	State	District	School	Norms	State	District	School	Norms	District	School
Q_3	111	110	105	121	7.2	7.3	6.9	9.7	7.0	6.8	8.5
Mdn	100	99	96	113	6.2	6.2	5.6	8.3	6.2	5.8	7.5
Q_1	89	89	85	101	4.9	4.7	4.1	6.9	5.1	4.4	6.8

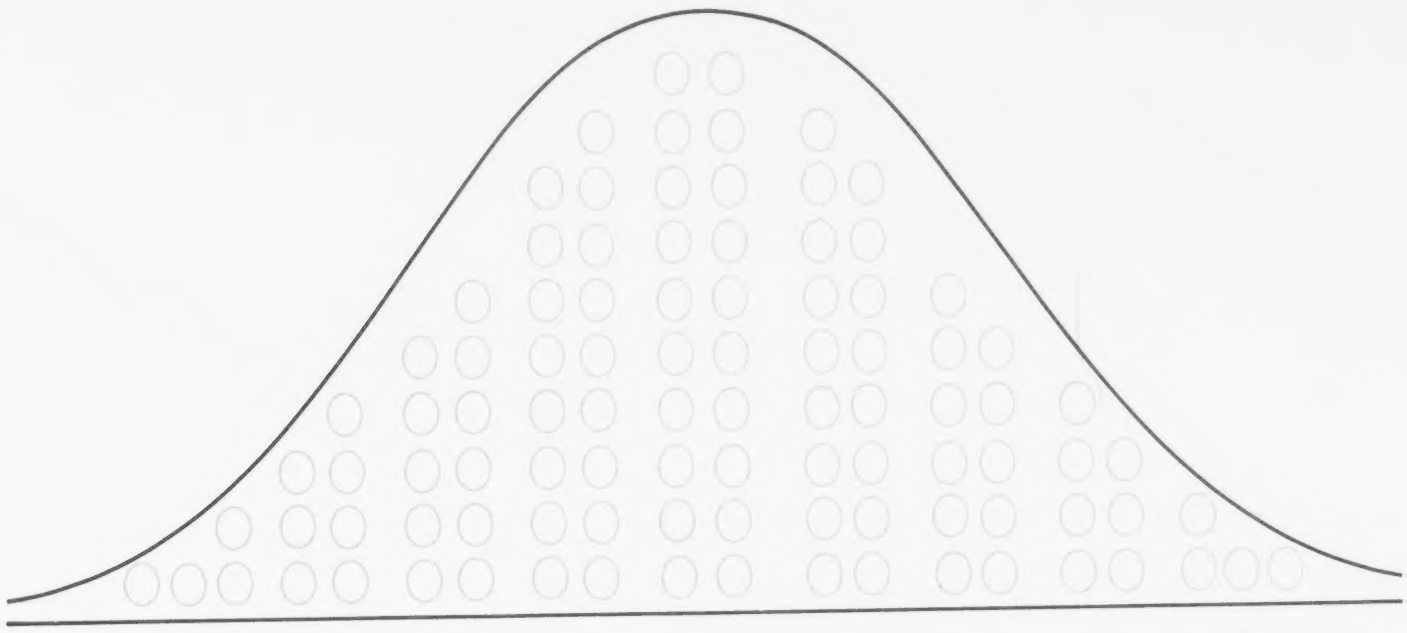
Figure 7. Results of 6th Grade Ability, Reading and Arithmetic Tests (Publisher's National Grade Equivalent Norms).

the normal skill levels, the percentages would correspond to the norms. Very few schools actually have this theoretical pattern, however.

This sort of look at scores provides an awareness that there are pupils in most schools performing at most of the skill levels. Also, it allows a closer look at how children of different abilities may be performing in the various achievement areas. Furthermore, any marked difference in the achievement level in different subject areas becomes visually apparent through obvious discrepancies in the number of students at the same level for different tests.

In making these various comparisons, however, there are a number of factors which should be considered in accounting for discrepancies among groups or across different types of tests. It is important to present a candid appraisal and include recognition of legitimate reasons why certain scores may be higher or lower than ordinarily anticipated. For example:

- Cultural bias in the test content may be a factor. In relation to the terminology and the types of questions asked, the test may not be completely satisfactory for a particular school population.



Norms % (Any Test)	4	7	12	17	20	17	12	7	4
SCAT	8	17	15	31	21	5	2	1	0
STEP MATH	10	17	19	22	20	9	2	1	0
STEP READING	8	9	26	23	17	8	5	3	1
STEP WRITING	4	7	18	36	14	12	5	3	0

Figure 8. Distribution and Range of Ability and Achievement Test Scores.

- A source of error which occurs at times with nationally standardized tests is that of basing the normative data upon an inadequate sample of students to estimate percentile and grade equivalent scores.

- Since test content, too, is based upon a sampling process, this factor can influence levels of test performance as well. Some tests may reflect accurately what is taught in a particular school; others may not.

- Unusual conditions in the administration of tests or the learning atmosphere of the school preceding the test can affect scores; although this may be true, the effect of such conditions is likely to be exaggerated in many instances.

- Emphasis in the curriculum can be a major source of differences in test scores across subject areas; e.g., if most of the mathematics instruction time is spent on number concepts, computation skills are likely to suffer. Another important variable along this line is the particular text used in a given course. Different texts will vary to the extent that they "match" different tests.

In conclusion, it is appropriate to remind parents that the test scores of their children are available to them by consulting with the teachers (or principal). It is generally preferable to use test results in an overall conference. Performance on various tests should be viewed along with how well a child is doing in class work, any special interests or problems which exist, and many other factors in order to have the greatest meaning. In other words, test scores are just one type of information which should be used in evaluating pupil progress.

TEST SCORES AND SCHOOL NEEDS

At some time during the discussion on test scores it may be appropriate to summarize and to discuss school needs as they relate to school achievement and progress. Questions can be solicited and discussions of this type may grow out of the questions asked. Citywide score patterns and those for many schools often emphasize the urgent educational needs of large numbers of pupils. Perhaps one of the values of release of school-by-school data is that of informing the public more fully about these needs, especially in the perspective of fiscal problems or other current community issues.

Among the general problems which might be brought into the discussion are:

A. Absenteeism — including excused and unexcused absences. These often constitute significant problems in a number of schools.

B. Physical well-being of pupils — including problems associated with nutrition, adequacy of clothing and shoes, and needs for treatment of major medical problems of pupils.

C. Environmental complement to school instruction — including the paramount importance of parent interest in the child's school activities, encouragement of reading at home, importance of discussing with and explaining to children (building language competence).

D. Pupil interest and motivation

E. Pupil mobility

F. Specific instructional problems at school related to budget restrictions — including materials and equipment, double sessions, etc.

It may be well to describe the school's efforts in attempting to meet some of the problems listed above. Communicating the feeling that the schools are also recognizing problems related to instruction (e.g., "This school needs to strengthen its arithmetic computation skills") as well as those attributed to the population it serves is essential to building confidence in the schools. Judiciously acknowledging weaknesses where they exist and indicating that efforts are being made to correct them is the strongest position that can be taken. It can be a sound position from which to make requests for needed parental and community support.

FREQUENTLY ASKED QUESTIONS

Certain questions tend to come up more often than others. Obviously, each must be handled in light of local circumstances. The following suggestions may be helpful, however, in anticipating difficult questions and in forming honest, helpful answers which take account of existing conditions in the school and community.

Is it true that test scores are related to family income?

Yes, as an overall statistical tendency. In other words, the score averages of low-income groups are lower than those for middle-income groups, and so forth. It is important to keep in mind, however, that these are tendencies and not hard and fast rules. There are many children from low-income families who are much brighter than the average child in the high-income family group. The reverse

9

is also true. There are some children in high-income families who are much slower than the average child in low-income families.

Do suburban districts give children a better education than city school districts? Are their test scores higher?

The answers to these two questions are "no" and "yes" respectively. Suburban school districts typically serve children from higher-income families. Therefore, the ability score averages do tend to be higher than the score averages of the cities. It is not surprising that their reading and arithmetic scores also tend to be higher. However, we generally find that children in our schools perform in reading just as well as children who receive similar scholastic aptitude test scores in other districts.

Aren't these tests biased against the disadvantaged or minority child?

The answer to this question depends largely on the uses which will be made of the results or the type of interpretations which are made or the kinds of conclusions that will be drawn. Certainly, scholastic ability tests, or so-called intelligence tests, are heavily weighted with language or verbal types of items. Children coming from surroundings where a different language or a very inadequate type of language is spoken will be at a disadvantage. It must be recognized that scores from verbal tests give only a minimal estimate of the true ability or potential of these children.

If these tests are used as the only factor in setting future expectancies or goals for any child, they are being misused. This is not to say that ability tests do not have value. While they do not measure true native intelligence, they give a score indicating how well-developed important verbal and reasoning skills are at the present time. They are useful in giving us an estimate of how well we can expect a child to be achieving in reading and other skill areas right now. These scores also give teachers and principals additional information regarding the type and pacing of instruction during the next few months.

While most children do tend to score at about the same scholastic ability level in subsequent testing, some children show impressive changes in scholastic ability just as they may also show improvements in spelling, reading, etc. It is, therefore, important to retest school ability at various grade levels.

A little different perspective should be taken with achievement tests. Regardless of a pupil's background, we are trying to determine how well he can perform certain tasks. How well can he add, subtract, multiply, and divide? How well can he read and interpret short paragraphs? How well does he spell? If a student is unable to perform these tasks in a satisfactory manner, we need to know it so that we can take corrective steps. If background factors are contributing to problems in achievement areas, they must be recognized; but the fact remains that these students are having difficulties in these areas, and we would like to help them improve their performance.

There is generally agreement that we want all pupils to be able to read well in the standard English language, to be able to express themselves in this language, and to be competent in many other basic skills areas. If bilingualism is causing a student to earn a low score on a reading test, we should recognize that the basic problem is a reading problem perhaps more than a testing problem.

Another key issue here relates to "norms" populations. When a test's national norms do not represent large cities where large numbers of disadvantaged children live, there is just reason for concern. State norms may provide some help in this area if all children in the state are included. In California, for example, we know that cities such as Oakland, San Francisco, San Diego, and Los Angeles are represented in the state comparison group.

Why should these scores become a part of a child's record? Won't they be used against him in later years?

In practically all school districts around the country, test scores are never released to anyone but the pupil, his parents, or a legal guardian, and, of course, interested school professionals such as teachers and counselors. In later years, when school records are needed for employment or entrance into college, only the transcript of grades for courses taken in grades 9-12 is requested by companies or colleges.

The school record, or cumulative folder, is a very valuable document. By accumulating a log of significant teacher judgments, health information, test scores, and other important information, teachers and principals are in a better position to make wise decisions and plans regarding the instructional program for each pupil. Perhaps a school record could be compared to a medical record and the uses which are made of it. When a doctor is evaluating a blood pressure reading, laboratory test, or other information, he does so in light of previous information on the chart. The questions may be, "Does this present blood pressure reading show a change for this individual? What treatments have been tried, and how did he respond?" Likewise, a teacher may ask questions such as:

"Is John's growth this year better or poorer than in previous years?"

10

"What are his particular interests and strengths?"

"In what areas does he need the most help?" etc.

Schools need a complete record of test scores in order to have a background of information for answering questions of this type.

How do district score averages compare with averages for other specific districts?

In some states the laws related to the state testing program include provisions making it illegal to publish listings of district-by-district averages. Therefore, this type of data may not be available for answering the above questions. However, in California legislation has been enacted not only allowing but also requiring the publication of district-by-district comparisons and rankings. Fortunately, efforts are also being made to provide additional information regarding the demographic characteristics of the districts. Overall, one generally finds that ability and achievement score averages vary right along with the socioeconomic levels of the districts.

It may be possible, by taking into consideration a number of critical socioeconomic and other school-influencing variables, to make some comparisons. These comparisons will still be very gross at best.

It may be of some use to lay audiences to know that large city school districts do tend to score lower than small suburban districts because there are parallel socioeconomic differences between these types of districts. When relevant factors are brought into consideration, it can generally be concluded that pupils in one district score at about the same level as pupils in other districts who have comparable ability levels and background advantages and/or disadvantages.

How do scores in this school compare to those in "X" school in the district?

One should avoid, as much as possible, extensive comparisons of one school with another unless it is done with considerable care. One should examine data on other pupil and community characteristics such as levels of family income, pupil mobility in and out of the school, ability levels and so forth.

With respect to an individual school, it may be useful to:

– bring out its general standing within the city; one can be rather specific here;

– mention comparison with other schools serving similar communities;

– point out that score distributions overlap considerably – schools with high averages still have pupils scoring well below national, state, and city averages; schools with low averages have some pupils scoring well above these national, state, and city averages. Discuss the issue of how well individuals are doing in relation to their scholastic ability test results. Comparisons can be made of medians and quartiles but should not overlook individual patterns of scores.

This may be the point to articulate specific problems the teachers are facing in relation to: double sessions, class size, instructional materials, pupil attendance, pupil motivation, etc.

It may be possible to indicate what continuing steps are being made to maximize achievement; to help each individual progress as rapidly as possible. The primary interest here is the individual child and his achievement. Improvements in test averages will result as we are successful.

Aren't we becoming overly concerned with tests? Will teachers start teaching for the test?

There is always a danger of losing perspective in regard to tests and test scores. The following points may be helpful:

- Test scores provide one important type of information in evaluating the progress of the pupils which should not be discounted. If we find that pupils are reading, on the average, below state and national norms, we must acknowledge these facts and make a thorough analysis of why this has happened.

- Tests are limited in that they do not cover a number of very important aspects of pupil growth – these include attitudes, appreciations, etc. Further, tests are sometimes limited by the fact that they may not adequately cover some major aspects of local curriculum.

- Tests must not determine what is taught in the schools. They must measure the outcomes of instruction which has been judged to be appropriate for a given grade level. However, if a test indicates that certain outcomes or goals are not being fulfilled, there is indication that a reevaluation of instruction should take place and perhaps reteaching in selected areas should be done. When the test happens to cover areas already designated for presentation at a grade level, there is no reason to avoid teaching the material even though it may be in the test. On the other hand, one should not teach a certain type of material just because it happens to be on the test.

11

SELECTED REFERENCES

Anderson, S.B., Katz, M., and Shimberg, B. *Meeting the Test.* New York: Scholastic Book Services, 1963.

Chauncey, H. and Dobbin, J.E. *Testing: Its Place in Education Today.* 1st ed. New York: Harper & Row Publishers, Inc., 1963.

Cronbach, L.J. *Essentials of Psychological Testing.* 2nd ed. New York: Harper & Bros., 1960.

Durost, W.N. How to tell parents about standardized test results. *Test Service Notebook,* No. 26. Tarrytown, N.Y.: Harcourt, Brace & World, Inc., 1961.

Durost, W.N. Why do we test your children? *Test Service Notebook,* No. 17. New York: Harcourt, Brace & World, Inc., 1956.

A glossary of measurement terms: ninety-six concepts which constitute a basic vocabulary in evaluation and testing. Del Monte Research Park, Monterey, Calif.: California Test Bureau, 1959.

Katz, M.R. (Ed.). *ETS Builds a Test.* Princeton, N.J.: Educational Testing Service, 1959.

Ricks, J.H., Jr. On telling parents about test results. *Test Service Bulletin,* No. 54. New York: The Psychological Corporation, 1959.

Thompson, A. Tentative guidelines for proper and improper practices with standardized achievement tests. Del Monte Research Park, Monterey, Calif.: California Test Bureau, 1959.

Traxler, A.E., Jacobs, R., Selover, M., and Townsend, A. *Introduction to Testing and the Use of Test Results in Public Schools.* New York: Harper & Bros., 1953.

Wesman, A.G. Aptitude, intelligence, and achievement. *Test Service Bulletin,* No. 51. New York: The Psychological Corporation, 1956.

NCME **NATIONAL COUNCIL ON MEASUREMENT IN EDUCATION**

Office of Evaluation Services
Michigan State University
East Lansing, Michigan 48823

Second class postage paid at
East Lansing, Michigan 48823

"PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL BY MICROFICHE ONLY HAS BEEN GRANTED BY

NCME

TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE U.S. OFFICE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER."

A series of special reports of the National Council on Measurement in Education